DOCUMENT RESUME

ED 326 101 HE 024 001

AUTHOR Smith, Dwight L.
TITLE Validity of Faculty Judgments of Student Performance: Relationship between Grades and Credits Earned and External Criterion Measures. ASHE Annual Meeting Paper.
PUB DATE Nov 90
NOTE 35p.; Paper presented at the Annual Meeting of the Association for the Study of Higher Education (Portland, OR, November 1-4, 1990).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Academic Achievement; *College Credits; College Faculty; *Evaluation Criteria; *Grades (Scholastic); Higher Education; Knowledge Level; Reliability; Standardized Tests; *Student Evaluation; Teacher Attitudes; Validity
IDENTIFIERS *ASHE Annual Meeting

ABSTRACT

A study investigated the validity and reliability of grades and credits earned by students in the five academic departments of biological sciences (33 majors), organic chemistry and biology (106 majors) at a large public university. Validity was assessed by comparing grades assigned by faculty with scores earned on subject-related standardized tests. Reliability was determined by statistical analysis of the stability of student grades across terms. Results showed consistently positive and often strong correlations between grades and student performance on external criterion measures in the field. These relationships are conservative estimates of the validity of grades. Evidence for the stability of student performance was also strong and consistent. The findings offer no support to claims that grades and credits earned are not valid or reliable measures of academic achievement. Includes 28 references. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED326101

Validity of Faculty Judgments of Student Performance: Relationship Between Grades and Credits Earned and External Criterion Measures

Dwight L. Smith

Blackburn College

Paper presented at the Association for the Study of Higher Education Conference

November 1, 1990

HE 024 001

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Dwight L. Smith

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ASHE

Texas A&M University
Department of Educational
Administration
College Station, TX 77843
(409) 845-0393

ASSOCIATION FOR THE STUDY OF HIGHER EDUCATION

This paper was presented at the annual meeting of the Association for the Study of Higher Education held at the Red Lion-Jantzen Beach in Portland, Oregon, November 1-4, 1990. This paper was reviewed by ASHE and was judged to be of high quality and of interest to others concerned with the research of higher education. It has therefore been selected to be included in the ERIC collection of ASHE conference papers.

Abstract

During the mid-1980s several national reports were critical of the traditional methods utilized by colleges and universities to determine student learning -- grades and credits. In this study, the validity and reliability of grades and credits earned of students in five academic departments at a large, public university were analyzed. Results show a positive an often strong correlation between grades assigned by faculty members and student performance on external, standardized measures of academic achievement. Findings for the validity of credits were not as supportive. There was generally strong and consistent evidence regarding the stability of student performance.

In national reports critical of the state of higher education (Association of American Colleges, 1986; Education Commission of the States, 1986; National Governors Association, 1986; and National Institute of Education, 1984), criticism has been leveled at the traditional measures colleges and universities have utilized to determine student academic achievement -- grades and accumulation of credits. The reports' authors take the position that other forms of assessment will provide better evidence of student learning. Moreover, it is their belief that, if implemented, these forms of assessment would provide evidence of student learning that public and governmental officials could then use to judge the institutions' worthiness (Fiske, 1987).

These positions on the assessment practices of colleges and universities are most curious considering the extent to which colleges and universities go to assess students. Boyer (1987) states this paradox most succinctly, "American undergraduates are forever being measured. They go to class, take notes, write papers, have pop quizzes, report orally, and receive interim marks and final grades in every course" (p. 251).

The discrepancy between the claims of the national reports and the practices of the colleges and universities is evident. The national reports claim that grades and credits are neither

valid nor reliable measures of student learning. Colleges and universities believe they are valid and reliable. And though the claims of the national reports have not been substantiated, it is at the same time the case that, unfortunately, most colleges and universities do not have evidence to contradict the opinions expressed in the national reports.

The purpose of this study was to investigate the extent to which the amount of formal study (i.e., number of credits earned in a discipline) and the institution's judgment about learning (i.e., grades earned) are reliable and valid indicators of student learning. The validity of credits and grades was investigated through a comparison of student performance on an external standardized criterion measure. The reliability of grades was examined by analyzing the stability of grades across courses within an academic department. The two forms of analyses should provide empirical evidence from which to consider the claims of the national reports.

Validity of Credits Earned

Studies concerning the assessment of academic achievement upon completion of undergraduate education provide evidence in support of the validity of credits earned. Chickering (1969) and Pace (1979) reported increases in academic achievement test

scores for students majoring in or taking more courses in a discipline than for non-majors of a particular discipline. Bowen (1977) and Pascarella (1985, 1989) reviewed studies which found changes not only in content knowledge for college students, but also in vocabulary and mathematical skills, critical thinking, intellectual tolerance, esthetic sensibility, and lifelong learning.

Despite evidence in the improvement in cognitive skills between freshman and senior years, there are limits to the generalization of these findings. First, there is inter- and intra-institutional variability (Chickering, 1969) meaning that these gains are not uniform between colleges or even within the same college. Between-college variation is particularly important, for some critics note that the weak performance of specific institutions is cause for concern. Second, much of the research concerning academic achievement occurred prior to the 1970s (Pace, 1979). Thus, findings indicating significant academic achievement among students in the 1950s and 1960s does little to address public concerns about academic achievement of students in the 1980s. Finally, the lack of systematic research offers "...spotty, partly obscure, and incomplete" (Bowen, 1977 p.97) evidence of undergraduate academic achievement.

Validity of Grades

Some of the current criticism about the use of grades and credits as measures of academic achievement is probably a result of the grade inflation phenomena of the 1960s and 1970s, a period during which the average grades in college climbed steadily with the grade of B becoming more common than the grade of C (Birnbaum, 1977).

In addition to the overall tendency in the rise of grades, there was an increased awareness of sizable differences in grading standards across the various disciplines (Goldman & Widawski, 1976; Strenta & Elliot, 1987; Warren, 1971; Willingham, 1985; Young, in press). Grading differences were reported between disciplines, with the students in the sciences receiving lower grades than students in the humanities and social sciences. However, pre-college academic achievement measures were higher for science majors than their counterparts in the humanities and social sciences. These findings strongly suggest that different standards are used between disciplines when assigning grades. If grading standards vary between disciplines to the point where students with stronger pre-college ability receive lower grades, then the meaning of grades can be called into question.

In the present discussion, these findings do not provide evidence of validity since the date from these studies were

collected sometime ago. The relationship between grades and outcomes measures has been suggested recently as way to investigate the validity of grades (NIE, 1984; Rutgers University, 1987). Hartnett and Schroder (1987) reported correlation coefficients between .36 and .56 for students' major grade point averages and scores on the respective GRE Subject Test for the academic discipline. Similar findings (Nowaczyck & Frey, 1982; Lollis, Einstein & Brewer, 1987) were reported for students majoring in psychology and scores on the Graduate Record Examination Psychology Test.

Findings such as these provide criterion-related evidence that grades are valid measures of academic achievement in the disciplines investigated. However, Messick's (1989) position on the need for several pieces of criterion-related evidence must be carefully weighed before the question of whether grades are valid measures is answered.

Reliability of Grades

The reliability of grades is also necessary to consider since the claims in the national reports would suggest that grades reflect a large degree of measurement error. If this is true, then grades should vary greatly between courses within a discipline. On the other hand, if grades can be considered

reliable, then they should be reasonably stable across courses in the discipline.

Warren's (1971) review of reliability studies of college grades indicated that freshmen grades across the curriculum had reliability coefficients of .70 - .84. However, when grades were considered across semesters the magnitude of the reliability coefficients decreased from the first to the eighth semester (Humphreys, 1968). The findings of Werts, Linn, and Joreskog (1978) which analyzed Humphreys' data supported his assertion that grades across the semester have a simplex form. Their analysis also supported the assumption that reliabilities across semesters were equal.

METHOD

Academic achievement among students at a large, public university was investigated by two academic departments, biological sciences and chemistry. These departments developed an assessment study utilizing an external criterion measure for student performance appropriate for the discipline. In each case, the selection of the test was made by faculty members in the department. For the biological sciences, students majoring in the department were tested with the Educational Testing Service Major Field Test for the Biological Sciences. In

chemistry, students who took a professional organic chemistry semester course were tested with the American Chemical Society professional certification exam for this subject matter. This study investigated the relationship between student performance on the external criterion measure and their grades and where applicable the number of credits.

The method used in determining the reliability of grades was the examination of the stability of students' grades across terms -- a method employed in previous studies. Two approaches were used to assess reliability. First, a correlation matrix of all possible pairs of required courses or exams was constructed. Significant correlation coefficients between course pairs would reflect consistency in course grades. Second, the consistency of grades across common courses was measured. This method considered whether the variation of grades received by students from course-to-course was less than the variation of all grades. If this were true, then the grades would be considered reliable. For the purpose of this study, a coefficient alpha which provides an estimate of variance on individual tests was employed with each course treated as a test item and all of the common courses the test. Using these two approaches, consistency of student performance would lead to the conclusion that grades are "reliable."

Since the two disciplines actively investigating academic achievement among majors were science disciplines, a social science and a humanities discipline were included in the study -- psychology and English. In addition, because the number of students participating in the biological sciences department's effort was small (thirty-eight), another biology department from the university was also included. There were two important methodological differences for these three departments. First, the external criterion measure used was the Graduate Record Examination Subject Test in biology, English, or psychology. The criterion measure was selected without consultation with the faculty of these departments, although the test was deemed appropriate since its content was determined by a national group of experts in the subject matter. Second, the students included in the study from these three departments were majors in one of these departments who voluntarily took the external criterion measure over a five year period from 1984 to 1989 and had their GRE scores reported to the university.

Finally, to replicate the studies cited by Pace (1979), a comparison was made between student scores on the external criterion measure and the number of credits earned within the academic discipline. This analysis was not performed for the chemistry department since that department involved students

within a particular course who had not completed their academic work in the major. The method for each academic discipline is outlined in Table 1.

Insert Table 1 about here

RESULTS

Validity of Grades and Credits Earned

Data summarizing the relationship between the overall grade point average (GPA), credits earned, and the respective external criterion measure are reported in Table 2.

Insert Table 2 about here

All correlations between GPA and the external criterion measures are positive and generally of such a magnitude to provide evidence that grades assigned by faculty members tend, in general, to be "valid" insofar as the various external tests are appropriate criteria. Correlations range from a .78 in chemistry to a low of .14 in English.

Positive relationships between number of credits earned and external criterion measures were found for the biological

sciences, English, and psychology departments. Note that for the purpose of this study only majors of the department were included. The majors had at the minimum 15 credits which limited the range of the number of credits earned and caused statistical limitations of the analysis of this relationship.

There was a very slight negative correlation (-.01) between the number of biology department credits earned and performance on the external criterion measure. It is possible that the GRE Biology Test is too general and fails to adequately differentiate between students' academic achievement. Horton (1959) in his review of the GRE Biology Test suggests just this; "In spite of the very broad nature of the knowledge and the high level of critical thinking, the test is not excessively specialized" (p. 836).

To summarize the date in Table 2, there were modest positive relationships found between GPA and the external criterion measure in the majority of the bi-variate relationships investigated. In the only department where there was no significant relationship found between the department GPA and the external criterion measure (English), this was most likely the result of the limited range of GPAs for the department's students (only three of thirty-five students had GPAs below a 3.00) in the study and the large differences between a criterion test of

based on writing. The chemistry department which included all students in a particular course, reported the highest correlation coefficient of the five departments (r=.78). This suggests that had a greater GPA range been the rule in the other departments, higher positive relationships probably would have been reported.

The relationship between department credits earned and the external criterion measure was not a conclusive, however. Three of the departments did have positive relationships between department credits and the score on the external criterion measure. It should be recalled from the studies cited by Pace (1979), that all students, no matter how few credits they earned (if any) in a particular discipline were included. Hence, if the present study had included freshmen and sophomores who were majors of the one of the disciplines included in the study, it is likely that the positive relationship between credits earned and scores on achievement tests reported by Pace would have been replicated. The limit in range of credits earned by students in the present study, thus, provided a conservative estimate for this relationship.

The second analysis of grade validity was the relationship between the external criterion subtests and the grade point averages in the subset of courses related to the content of the respective subtest. For the biology and biological sciences the

respective subtest. For the biology and biological sciences the content of the subtests were identical even though one criterion measure was the GRE Subject Test for the former and the ETS Major Filed Examination for the latter. The correlation coefficients between the three subtests and GPA are reported in Table 3.

Insert Table 3 about here

In all three disciplines there is a positive relationship between student performance on the subtests and the student's GPA in the appropriate subset of courses within the discipline. The magnitude of the correlations range from a high of .48 (cellular and subcellular subtest in the biological sciences) to a low of .12 (social psychology). Note that five of the eight correlations reported in Table 3 are .36 or above.

Overall, for the three departments there was a significant relationship found between a subtest and the GPA for courses related to the content of the subtest. For biology and the biological sciences there was a significant relationship between GPA and two of the three subtests and psychology GPA was found to be significantly related to one of the subtests.

Reliability of Grades

The reliability of grades was investigated in two ways. First, correlation coefficients were calculated for all possible pairs of courses which were required of majors in the department. In the case of the chemistry department, correlation coefficients were calculated for all possible pairs of tests and recitation scores. High correlations between courses, indicate student performance through the set of required courses or in the chemistry department the tests.

The summary of reliability coefficients between pairs of common courses in the four academic departments and in the case of chemistry, common tests and the recitation, is reported along with the coefficient alpha for these common courses in Table 4.

Insert Table 4 about here

Positive relationships were reported for the majority of pairs of courses with moderate-to-strong median correlations for pairs of courses in the disciplines. There were negative relationships found in two disciplines, biological sciences and psychology. (Correlation coefficients for pairs of courses in the respective departments are available upon request.)

The final measure of reliability for each department was the calculation of the coefficient alpha for the grades in the common courses. As mentioned previously, each course was treated as an item in a test with all the courses treated as the test. In this manner, the variation of grades across courses was compared to the variation of grades within a course. If this comparison provided evidence that the variation of grades across courses was smaller than the variation of grades in a course, then grades would be reliable. All of the five coefficient alphas in Table 4 are high positive values. This suggests that in common courses and, in the case of chemistry, tests student performance is stable and reliable.

The reliability evidence does reveal that departmental grades when considered in the aggregate as was done with the coefficient alpha tended to be reliable. The relationship between pairs of courses also indicated the reliability of grades even though negative or low positive relationships would not normally be construed as such. In the case of psychology, these relationships were found between courses whose content (Quantitative Methods and Social Psychology) demanded different types of cognitive abilities (verbal vs. mathematical) and differences in types of behavior graded (writing vs. computational). Therefore, differences in grades is more readily

apparent when analyzing pairs of courses. These results when taken together do provide evidence that grades tended to be stable from course-to-course.

Construct Validity of External Criterion Measures

A final analysis was conducted for the intercorrelational relationship between the GRE, SAT total scores, the GRE Subject Test or in the case of the biological sciences department the ETS Major Field Examination and departmental GPA. Pike (1989) raised concerns about the use of external tests such as the ACT College Outcomes Measurement Program in the measurement of program quality and suggested that the validity of such measures be investigated. The purpose of this analysis was to determine whether the external criterion measures used in the study were appropriate measures of academic achievement for the respective disciplines. If they are appropriate, then the relationship between GPA and the external criterion measure would be stronger than the relationship between the external criterion measure and the two tests of general academic ability, SAT and GRE General Test. However, if the relationship between the external criterion and the two tests of general academic ability were stronger, then there is a question of whether the external criterion measure is an appropriate measure of academic

achievement for the respective discipline. This latter situation would suggest all three tests are measures of general academic ability and that grades provide better evidence of academic achievement within the four disciplines. These intercorrelations are reported in Table 5.

Insert Table 5 about here

As indicated in Table 5 there is a stronger relationship in three of the four disciplines (biology, English, and psychology) between the three tests (SAT, GRE Subject Test, and GRE) than for the relationship between GPA and the GRE Subject Test. The only discipline which had a stronger relationship between GPA and the achievement test was biological sciences in which the faculty selected the external criterion measure. The strong interrelationship between the three tests in biology, English, and psychology does call into question the appropriateness of the use of the GRE Subject Test as an external criterion measure. Instead of measuring academic achievement for the discipline, the GRE Subject Test appears to replicate the measure of general academic ability of the students. Horton (1959) reported similar results, "The extent to which this test [GRE Biology Subject Test] requires a high level of critical ability may well account

for the fact that scores on it correlate .74 with score on the verbal section and .64 with scores on the quantitative section of the GRE Aptitude Test" (p.810). Therefore, to the extent that ability to perform well on tests of general academic ability is a key component of achievement test performance, suggests that grades, not achievement tests are the better indication of student performance. Grades, particularly in English and psychology, represent measures of cognitive abilities not readily measured by "right or wrong" answer tests.

Recall also that there were low and negative relationships between credits earned and the GRE Subject Test in biology ($r = -.01$), English ($r = .01$), and psychology ($r = .23$). The high relationship in each of these disciplines between the SAT, GRE Subject Test, and the GRE strongly suggests that the GRE Subject Test was measuring general academic ability instead of academic achievement in each discipline. Thus, a student with as few as fifteen credits earned was able to perform as well as the students with a greater number of credits earned on the external criterion measure. In the case of the biological sciences where faculty selected a standardized, external criterion measure, there was moderate relationship between credits earned and scores on the external criterion measure. ($r = .42$). These results suggest that credits earned provide better evidence of the

breadth and understanding of the subject matter studied than the GRE Subject Test.

DISCUSSION

The claims made by several national reports that grades and credits earned are not valid nor reliable measures of academic achievement were not supported by the findings of this study. In all five departments -- departments in the sciences, social sciences, and humanities -- there were consistently positive and often strong correlations between grades assigned by faculty members and student performance on an external, standardized measure of academic achievement in the field. Furthermore, for reasons cited earlier, these reported relationships are very conservative estimates of the validity of college grades. Thus, at one large, public university, it can be concluded that the grades received by students are valid measures of academic achievement.

In the one case where a wide range of grades was analyzed (chemistry, N=599), the correlation coefficient was .78 and significant at the .0001 level. If the other departments had included a greater range of grades as was the case with the chemistry department, the actual grade - external criterion measure relationship probably would have been much stronger.

Thus, these findings which were conservative estimates provide strong evidence for the validity of grades as measures of academic.

The findings for the validity of credits were not as supportive and in one case contradictory to those reported by Pace (1979). One would expect moderate-to-strong positive correlations, for as Pace (1979) argued on the evidence he reviewed, "Students learn what they study and the more the study, the more they learn" (p. 18). There is a significant difference, however, between this study and the ones reported by Pace. This study included only those students majoring in the academic department, whereas the study cited by Pace included not only majors, but all students taking a statewide test. This suggests that if students who had few credits in the academic department had taken the external criterion test, there may have been stronger relationships reported.

More importantly, however, was that the tests do not appear to have been "sensitive" to the breadth of student learning. Instead the external criterion test were measures of general academic knowledge and failed to differentiate students' academic achievement within a discipline. This shortcoming of the GRE Subject Test as a criterion measure for credits earned gives

further credence to Pace's claim of credits as being valid measures of academic achievement within a discipline.

There was generally strong and consistent evidence regarding the stability of student performance. Median correlations between grades in pairs of courses were above .40 in all the disciplines and there were strong positive coefficient alpha values for common courses in each of the disciplines. Feldt and Brennan (1989) criticized traditional notions of reliability standards as being unrealistic and "over inflated" because such standards failed to take into account performance over an extended period of time. This study did measure student performance over an extended period of time. The fact that there were moderate relationships reported for the majority of pairs of common courses suggests that there was consistency in student performance, and, in turn the grades received.

It should be noted that this study focused on criterion-related evidence in the determination of validity and reliability of grades and credits earned. Yet as Messick (1989) has pointed out, criterion- and content-related evidence are difficult to differentiate and evidence for one type of validity needs to be supported by the other. Thus, there is a need to investigate content-related validity evidence for grades and credits earned.

Nonetheless, the idea offered by the previously cited national reports that grades and credits earned are inadequate measures of academic achievement is not supported by the results of the study. In fact, the evidence does support the validity and reliability of these traditional measures of academic achievement at one large, public university. Hence, concerns about the traditional measures of academic achievement used by colleges and universities may not be warranted, especially in regards to these five academic departments. A bigger concern may be with the types of criterion measures used to measure academic achievement in these five departments.

References

Association of American Colleges. (1986). Integrity in the college curriculum: A report to the academic community: The findings and recommendations of the project on redefining the meaning and purpose of baccalaureate degrees. Washington, D.C.: American Council on Education.

Birnbaum, R. (1977). Factors related to university grade inflation. Journal of Higher Education, 48(5), 519-539.

Bowen, H. (1977). Investment in learning. San Francisco: Jossey-Bass.

Boyer, E. (1987). College: The undergraduate experience in America. New York: Harper and Row.

Chickering, A.W. (1969). Education and identity. San Francisco: Jossey-Bass.

Education Commission of the States. (1986). Transforming the state role in undergraduate education. Denver, CO: Education of the States.

Feldt, L.S. and Brennan, R.L. (1989). Reliability. In R.L. Linn (Ed.), Educational Measurement (3rd ed.) (pp. 105-146). New York: Macmillan Publishing Co.

Fiske, E. (1987, January 10). Colleges prodded to prove worth. New York Times, pp. 1, 27.

Goldman, R.D. and Widawske, M.H. (1976). A with-in subjects technique for comparing college grading standards: Implications in the validity of the evaluation of college achievement. Educational and Psychological Measurement, 36, 381-390.

Hartnett, R.T. and Schroder, B.C. (1987). The validity of undergraduate grades. Unpublished manuscript, Rutgers, The State University of New Jersey, Office of Educational Policy Studies, New Brunswick.

Horton, C.W. (1959). The Graduate Record Examination Advanced Tests: Biology. In O.K. Buros (Ed.), The Fifth Mental Measurements Yearbook (pp. 810-811). Highland Park, NJ: The Gryphon Press.

Humphreys, L.G. (1968). The fleeting nature of college academic success. Journal of Educational Psychology, 59, 375-380.

Lollis, T.J., Einstein, G.O., and Brewer, C.L. (1987). Using undergraduate factors to predict psychology Graduate Record Examination scores. Teaching of Psychology, 14(4), 202-206.

Messick, S. (1989). Validity. In R.L. Linn (Ed.), Educational Measurement (3rd ed.) (pp. 13-103). New York: Macmillan Publishing Co.

National Governors Association. (1986). Time for results: The governors' 1991 report on education. Task Force on College Quality. Washington, D.C.: National Governors Association.

National Institute of Education. (1984). Involvement in learning: Realizing the potential of American higher education. Washington, D.C.: National Institute of Education.

Nowaczyck, R.H. & Frey, J.D. (1982). Factors related to performance on the GRE Advanced Psychology Test. Teaching of Psychology, 9, 163-165.

Pace, C.R. (1979). Measuring outcomes of college: Fifty years of findings and recommendations for the future. San Francisco: Jossey-Bass.

Pascarella, E.T. (1989). The development of critical thinking: Does college make a difference? Journal of College Student Development, 30, 19-26.

Pascarella, E.T. (1985). College environmental influences on learning and cognitive development: A critical review and synthesis. In J.Smart (Ed.), Higher Education: Handbook of theory and research (pp. 1-61). New York: Agathon.

Pike, G.R. (1989). Background, college experiences, and ACT-COMP Exam: Using construct validity to evaluate assessment instruments. The Review of Higher Education, 13(1), 91-117.

Rutgers, The State University of New Jersey. (1987). Interim report of the University Academic Forum Committee on the Specification and Measurement of Academic Outcomes. Unpublished Manuscript, Rutgers, The State University of New Jersey, New Brunswick.

Strenta, A.C. and Elliot, R. (1987). Differential grading standards revisited. Journal of Educational Measurement, 24(4), 281-291.

Warren, J.R. (1971). College grading practices: An overview. Princeton, NJ: Educational Testing Services.

Werts, C., Linn, R.L., and Joreskog, K.G. (1978). Reliability of college grades from longitudinal data. Educational and Psychological Measurement, 38, 89-95.

Willingham, W.W. (1985). Success in college. New York: College Entrance Examination Board.

Willingham, W.W. (1974). Predicting success in graduate education. Science. 183, 273-278.

Young, J. W. (In press). Adjusting the cumulative GPA using Item Response Theory. Journal of Educational Measurement.

Table 1

Summary of Programs, Measures and Methods

Discipline	Students	Criteria	Reliability Procedure
Biological Sciences	Majors (N=33)	ETS Major Field Exam	Course pair correlations and alpha coefficient
Biology	Majors (N=106)	GRE Subject Test	Course pair correlations and alpha coefficient
Chemistry	Students in Organic Chemistry (N=599)	American Chemical Society Organic Chemistry Exam	Tests correlations and alpha coefficient
English	Majors (N=35)	GRE Subject Test	Course pair correlations and alpha coefficient
Psychology	Majors (N=112)	GRE Subject Test	Course pair correlations and alpha coefficient

Table 2

Correlations Between Faculty Judgments (Grades, Test Scores, and Credits Earned) and External Criterion Measures

	GPA and Test Score Correlations	Credits Earned Correlations	n
Department			
Biology	.42**	-.01	106
Biological Sciences	.59*	.41*	33
Chemistry	.78**	n/a	599
English	.14	.01	35
Psychology	.39**	.23*	112

NOTE: External Criterion Measures for Biology, English and Psychology GRE Subject Test; For Biological Sciences ETS Major Field Test; For Chemistry American Chemical Society Examination

**p<.001 *p<.01

Table 3

Correlations Between Student Performance on Various Examination Tests and Departmental GPA in Biology, Biological Sciences, and Psychology

Sub-test	Correlation with GPA in appropriate courses	n
Biology		
Cellular and Subcellular	.40**	104
Organismal	.20	90
Population, Ecology and Evolution	.48*	21
Biological Sciences		
Cellular and Subcellular	.48*	19
Organismal	.40*	33
Population, Ecology and Evolution	.15	28
Psychology		
Experimental	.36**	108
Social	.12	80

**p<.01 *p<.05

Note: Sub-test scores are in some instances based on the ETS Major Field Exams and in other instances on the GRE Subject Tests. Refer to Table 1 for details.

Table 4

Correlations Between and Coefficient Alpha For Grades in Common Courses, by Discipline

	Coefficient Alpha	Median r	Low r	High r
Biology: General Biology I, General biology II, physiology, and genetics	.89	.42**	.30	.56**
Biological Sciences: General biology I, General biology II, genetics, organismal biology, and ecology	.80	.42*	-.23	.68**
Chemistry: Four tests, one recitation	.82*	.49**	.37**	.61**
English: Expository Writing, LiteraryStudy I, Literary Study II, Shakespeare	1.06**	.57*	.38	.64**
Psychology: General, quantitative methods, social, personality, abnormal, learning, learning laboratory, child development, psychotherapy systems	.93**	.32*	-.07	.67**

**p<.01 *p<.05

Note: Reliability coefficients are based on different numbers of intra-course correlations, as indicated in the discipline lead-in. Thus, there were six possible intercorrelations in biology, where there were four common courses (n-1+n-2+n-3....+n-n) and 36 inter-correlations in psychology, where there were nine common courses.

Table 5

Intercorrelations Between Pre-college Ability measure (SAT), College Grades in Major Field (GPA), External Standardized Achievement Measure (GRE Field or Subject Test), and GRE General Test for Four Disciplines

Biology:	(1)	(2)	(3)	(4)
(1) SAT	---			
(2) GPA	.38** (85)	---		
(3) GRE Subject	.51** (85)	.42** (106)	---	
(4) GRE	.80** (74)	.44** (93)	.64** (93)	---

Biological Sciences:	(1)	(2)	(3)	(4)
(1) SAT	---			
(2) GPA	.09 (25)	---		
(3) GRE Field	.34 (25)	.59** (33)	---	
(4) GRE	----------			

English:	(1)	(2)	(3)	(4)
(1) SAT	---			
(2) GPA	.05 (32)	---		
(3) GRE Subject	.59** (32)	.14 (35)	---	
(4) GRE	.91** (28)	.06 (31)	.61** (31)	---

Psychology:	(1)	(2)	(3)	(4)
(1) SAT	---			
(2) GPA	.19 (89)	---		
(3) GRE Subject	.47** (89)	.39** (112)	---	
(4) GRE	.78** (78)	.48** (94)	.56** (94)	---

**p<.01

Note: (n) Underlined correlation coefficients are validities also reported in Table 2.